Runs Determined in a Sample by an Arbitrary Cut

By PAUL S. OLMSTEAD

(Manuscript received August 9, 1957)

This paper, after making a critical review of the literature pertaining to runs above and below in a fixed sample, provides the following extensions:

1. *Sample arrangement distributions for runs of length at least s on one, each, and either side of* any selected *cut for samples of 10 and 20,*
2. *Sample arrangement distributions for runs of length at least s on one, each, and either side of the* median *for samples of 10, 20, 40, 60, 100, and 200,*
3. *Sample arrangement distributions for runs of length at least s on each side of* all possible *cuts for samples of 10, 20, 40, and 100,*
4. *Asymptotic values of the probabilities of such arrangements when the sample size and length of run are large,*
5. *Convenient charts and tables for probabilities of 0.01, 0.10, 0.50, 0.90, and 0.99 to facilitate use by engineers and scientists, and*
6. *Discussion of a simple application.*

The inclusion of the case for runs of length at least s on each side of all possible *cuts should prove very useful because it provides a quantitative measure for a common operational procedure for which the exact probabilities were heretofore unknown.*

I. SUMMARY

This paper discusses certain nonparametic measures for use in detecting the presence of assignable causes in experimental data. Specifically, it assumes that a sample of n observations of a characteristic, X, has been obtained and that a particular arrangement, $X_1, X_2, \cdots X_n$, e.g., by the time order of determination or other considerations, increases the value of the sample as evidence. Assuming a cut at a particular value of X, such as A, such a series may be divided into groups of consecutive observations that lie, alternately, above and below the cut. The length of such a group is called a run.

The paper also presents charts, tables, and formulas relating to such sample arrangement distributions for runs above and below *any selected* and *all possible* cuts or demarcation values. Specifically, it contains:

a. A review of the literature relating to runs above and below (Section II).

b. Appropriate charts and tables for the convenience of the engineer or other user (Section II and III).

c. An example (Section III) and reference to others (Section II).

d. A procedure for obtaining the probability that a randomly selected arrangement of a sample of size n will contain one or more runs of length at least s *on each side* of at least one of *all possible* cuts or demarcation values that do not coincide with one of the numerical values in the sample (Section VI).

e. Relationships between n and s for constant probability (Section VIII).

f. The probability that a randomly selected arrangement of a sample of size n will contain one or more runs of length at least s on each side of a selected cut or demarcation value such that n_1 numerical values are above and n_2 numerical values are below ($n = n_1 + n_2$). Similar probabilities are given for arrangements with runs *above*, with runs *below* and with runs *on either side* of such a cut or demarcation value (Section IV).

g. Simplified formulas for runs above and below the median that are equivalent to those given by Mosteller[4] (Section V).

h. Asymptotic values of these probabilities for both n and s large (Section VII).

II. HISTORICAL BACKGROUND AND DISCUSSION

Runs above and below the average, the median, or some other selected value have been used by a number of engineers to assist in detecting and identifying assignable causes of variation in connection with research and development work. In order to have a clear picture of the problems of such work, it may be worthwhile to set down some statements which characterize it:

a. A repetitive process that has not been examined for control by statistical methods and that has not subsequently been brought into control is very unlikely to be in statistical control,

b. Causes of lack of control often occur sporadically, being present for relatively short intervals of time,

c. Such causes of lack of control may often be detected by taking account of order either in manufacture or in taking observations, and

d. A basis for determining what fractions or portions of the observations may have been affected by an undesired cause is the application of statistical tests to the pattern of the individual values of the measurements in the order in which they were obtained.

Runs above and below have been particularly useful in assisting in the identification of such assignable causes. Their use in engineering has progressed through the following steps:

1. Using a procedure based on the work of Cochran,[2] Shewhart[5] showed the distribution witb respect to length of the runs above and below the average. It was his observation that a run of length 7 was often associated with a cause that could be found. Cochran had derived the distribution of runs of lengths s (our notation) of two complementary events E_1 and E_2 of known probability, p, and $q = 1 - p$, respectively. In applying Cochran's formula, Shewhart chose two statistics, $\bar{X}$ and p, from his observed data. Recognizing that this might invalidate the use of Cochran's formula, he suggested to the writer that this loophole could be avoided by working out the distribution for run lengths relative to the median. This distribution was worked out and recorded in a memorandum dated October 14, 1940.

2. About the same time, Mood[3] was working on his "Distribution Theory of Runs" for which the distribution relative to the median is a special case. He included in his results expressions for the variances and covariances. Campbell[1] made use of the distribution of lengths of run relative to the median.

3. The next step was to obtain the distribution of possible arrangements with runs of at least a given length relative to the median. Mood[3] gave a general analysis of the problem, which was supplemented in a form more easily comprehensible to the engineer by Mosteller.[4] Mosteller gave criteria based on sample size at given probability levels for length of run on one side and on either side of tbe median. While this paper was in preparation, Olmstead bad been examining the problem of the probability of arrangements with runs of at least a given length on each side of the median. When this was brought to Mosteller's attention, his paper was revised to include this case which had its inception in the engineering idea that if two cause systems were operating in separate periods they would be likely to produce separate groups of high and low values.

4. Following this, attention was given to the distribution of arrangements, as indicated in Section V of this paper, where division for runs above and below was made at some location other than the median. Validity in use of the probabilities calculated on this basis was dependent on the choice of division location prior to the test and often left the en-

gineer and the statistician uncertain concerning the risks that were being taken wben the division location was chosen after looking at tbe data. Because of assumption (a) above, this did not worry the engineer as much as it did tbe statistician, particularly when the engineer could find a cause associated with long runs identified in tbis way. The fact that he usually found such a cause indicated that some other way of considering the problem from the viewpoint of mathematical statistics would be fruitful.

5. The obvious next step was to find a procedure for counting all of the possible arrangements of n numbers, no two alike, that would have one or more runs of length at least s on each side of at least one of all of the possible division points that do not coincide with one of the numerical values in the sample. One way of doing this is first to write down or plot all ($n!$) possible arrangements of the n numbers. Assume tbat tbe numerical values of the numbers are the y-coordinates and the order in which they occur in an arrangement is indicated by the x-coordinates of such a plot. All such plots could then be examined to see what y-division not at one of the y-values would give tbe longest run of consecutive y-values on each side of the division. In tbis way, each arrangement would be assigned to a category where a particular length of run was equalled or exceeded on each side for at least one of the possible y-divi-

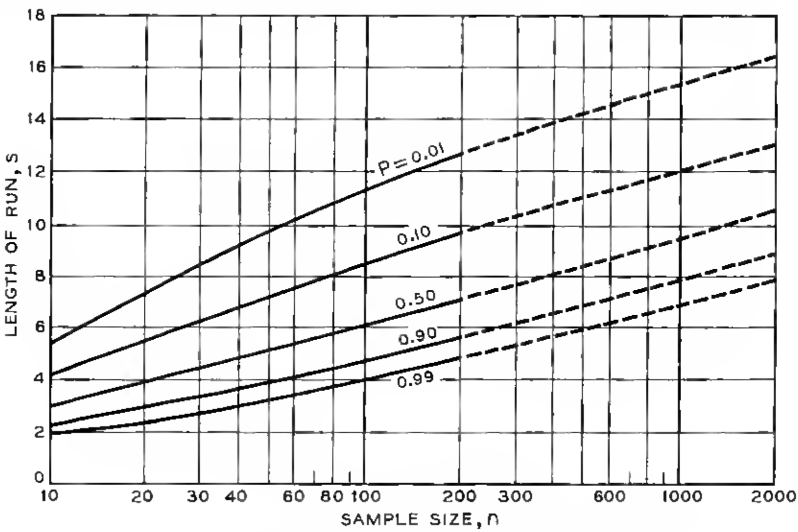

Fig. 1 — Length, s, of run on *one* side of *median* versus sample size, n, for selected values of probability, P.

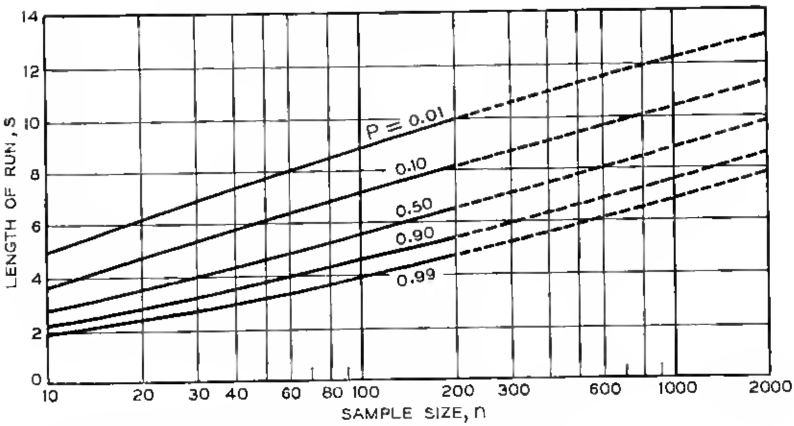

FIG. 2 — Length, s, of run on *each* side of *median* versus sample size, n, for selected values of probability, P.

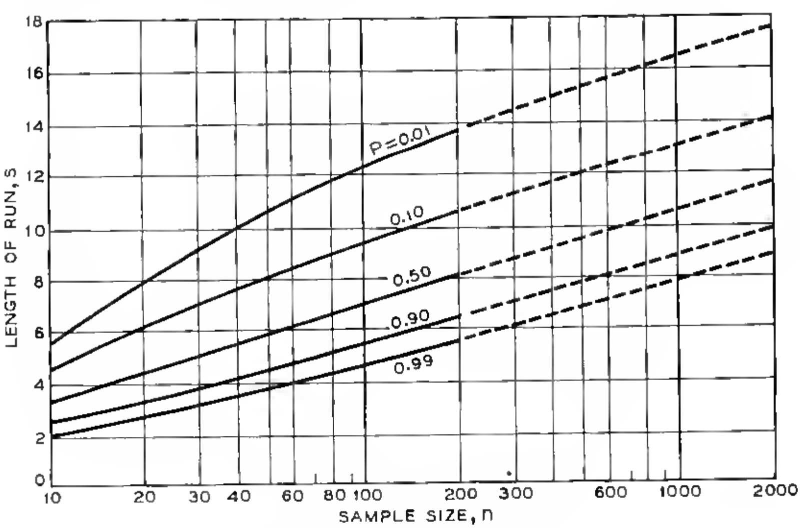

FIG. 3 — Length, s, of run on *either* side of *median* versus sample size, n, for selected values of probability, P.

sions. The process presented in Section VI is the mathematical equivalent of carrying out such a count. This process is gratifying to the engineer and the statistician alike because of the freedom permitted in setting the division location after examining the data so as to obtain the longest lengths of run on each side of the selected value. Use of this information

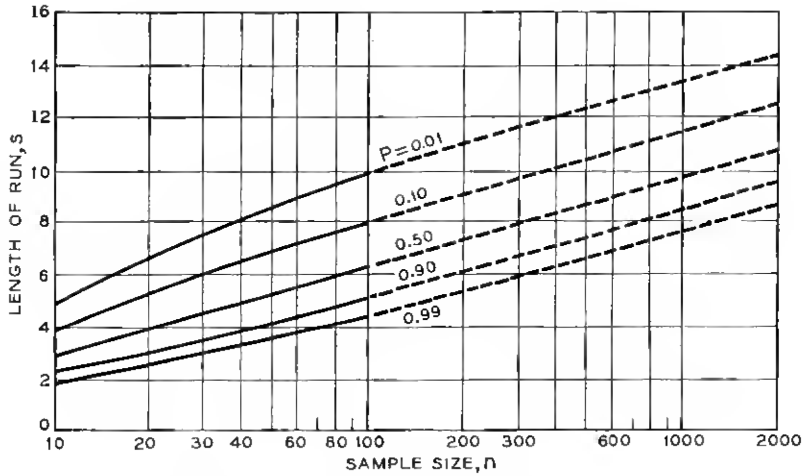

FIG. 4 — Length, s, of run on *each* side for *any cut* versus sample size, n, for selected values of probability, P.

was made first in an article by Walker and Olmstead.[6] Its part in detecting the type of an assignable cause appeared first in an article by Olmstead.[7]

6. In connection with the investigation undertaken for this paper, the asymptotic relationships for determining probabilities when n and s are large have been obtained (Section VII) and the results compared with those given by the exact relationships. The exact relationships applying to the median have been calculated for sample sizes of 60, 100, and 200 extending this information beyond the range usually covered by research workers. For the convenience of such workers, four charts (Figs. 1, 2, 3, and 4) have been prepared to show the relationships between s and n for $P = 0.01$, 0.10, 0.50, 0.90, and 0.99 for the primary types of runs.

III. WORKING TECHNIQUES

As just mentioned, Figs. 1, 2, 3, and 4 present graphically five percentage points of each of the four "above" and (or) "below" run distributions for all sample sizes from 10 to 2,000. The same information is furnished in tabular form in Tables I, II, III, and IV. How these are derived and calculated is discussed later (Sections V, VI, and VIII). Specifically, the four types of distribution thus made available are:

a. The probability, P, of the event that the length of the longest run on one pre-chosen side of median equals or exceeds s; if above, the prob-

ability is designated $P(s/-, \text{medium})$; if below, $P(-/s, \text{median})$. The notation, $P(s/-, \text{median})$ may be read — the probability that an arrangement will contain a run of length at least s above the median.

b. The probability of the event that the length of the shorter of the longest run above and the longest run below the median equals or exceeds s: designated $P(s/s, \text{median})$, where s/s means that there is a run of length at least s, on each side of the median.

c. The probability, P, of the event that the length of the longer of the longest run above and the longest run below the median equals or exceeds s: designated $P(s, \text{median})$, where s means the longer of $(s/-$ median) and $(-/s$, median).

d. The probability, P, of the event that the length of the shorter of

TABLE I

Minimum sample sizes, n, that exceed selected probabilities, P, for a given length, s, of run on *one* side of median calculated from Table XVI and equations (23) and (27) to three significant figures.

Run Length s	Probability, P				
	0.01	0.10	0.50	0.90	0.99
1	2	2	2	2	2
2	4	4	6	8	12
3	6	6	12	22	38
4	8	10	22	54	100
5	10	16	46	116	230
6	14	26	92	260	490
7	18	44	182	530	1044
8	26	78	360	1104	2140
9	38	142	714	2240	4370
10	56	256	1424	4530	8980
11	86	480	2850	9190	18240
12	140	930	5680	18540	37200
13	234	1838	11330	37600	75500
14	410	3630	22700	75700	151700
15	748	7160	45300	151700	303000
16	1446	14190	90600	303000	607000
17	2830	28100	181200	607000	1214000
18	5530	56100	362000	1214000	2430000
19	10860	117300	725000	2430000	4850000
20	21500	235000	1450000	4850000	9710000

Examples of use:

	Observed Data		*Probability, P*
Case 1	$n = 96$	$s = 4$	$0.90 < P < 0.99$
2	54	10	$P < 0.01$
3	56	10	$0.01 < P < 0.10$

TABLE II

Minimum sample sizes, n, that exceed selected probabilities, P, for a given length, s, of run on each side of median calculated from Table XVI and equations (24) and (27) to three significant figures.

Run Length s	Probability, P				
	0.01	0.10	0.50	0.90	0.99
1	2	2	2	2	2
2	4	4	6	10	14
3	6	8	14	26	44
4	8	14	30	68	116
5	12	26	68	152	252
6	20	50	140	322	552
7	34	98	290	676	1164
8	62	194	596	1390	2390
9	116	390	1208	2830	4930
10	216	782	2440	5650	10140
11	446	1182	4910	11750	20700
12	884	2360	9840	23800	42500
13	1762	4720	19890	48600	86700
14	3510	9450	39900	98600	174200
15	6990	18900	80500	197300	348000
16	13930	37800	161300	395000	697000
17	27900	75600	323000	789000	1394000
18	55500	151200	645000	1578000	2790000
19	111000	302000	1290000	3160000	5570000
20	222000	605000	2580000	6310000	11150000

TABLE III

Minimum sample sizes, n, that exceed selected probabilities, P, for a given length, s, of run on either side of median calculated from Table XVI and equations (25) and (27) to three significant figures.

Run Length s	Probability, P				
	0.01	0.10	0.50	0.90	0.99
1	2	2	2	2	2
2	4	4	4	8	10
3	6	6	8	16	28
4	8	8	16	36	64
5	10	14	30	76	136
6	12	20	58	152	282
7	16	32	106	296	568
8	22	52	200	580	1150
9	32	86	388	1174	2310
10	42	150	758	2350	4640
11	62	262	1488	4720	9330
12	94	500	2920	9460	18730
13	156	962	5860	10660	37700
14	256	1876	11250	21300	75700
15	418	3670	22600	42600	151600
16	766	7330	45200	85300	303000
17	1472	14090	90100	170500	606000
18	2860	27900	180300	341000	1213000
19	5570	55500	361000	682000	2430000
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

TABLE IV

Minimum sample sizes, n, that exceed selected probabilities, P, for a given length, s, of run on each side of any cut calculated from Table XVI and equations (26) and (27) to three significant figures.

Run Length s	Probability, P				
	0.01	0.10	0.50	0.90	0.99
1	2	2	2	2	2
2	4	4	6	8	12
3	6	8	12	22	34
4	8	12	22	48	76
5	12	18	46	96	162
6	16	34	86	192	380
7	24	58	166	382	668
8	38	108	324	760	1342
9	66	204	638	1518	2690
10	118	400	1266	3030	5410
11	228	790	2530	6070	10870
12	444	1568	5050	12130	21500
13	878	3130	10070	24300	43100
14	1750	6220	20100	48500	86200
15	3480	12490	40300	97000	172300
16	6790	25000	80600	194100	345000
17	13860	49900	161100	388000	689000
18	27700	99900	322000	776000	1379000
19	55400	199800	644000	1553000	2760000
20	110800	400000	1289000	3110000	5510000

TABLE V

Speedometer readings at one minute intervals.

Time	MPH	Time	MPH	Time	MPH	Time	MPH
1	48	15	55	29	52	43	60
2	50	16	53	30	58	44	58
3	48	17	48	31	55	45	55
4	50	18	50	32	57	46	57
5	52	19	50	33	58	47	57
6	49	20	55	34	58	48	53
7	50	21	55	35	58	49	57
8	47	22	55	36	58	50	58
9	51	23	55	37	58	51	58
10	50	24	55	38	58	52	56
11	49	25	51	39	55	53	58
12	52	26	53	40	56	54	63
13	53	27	52	41	57	55	60
14	53	28	51	42	56	56	50

the longest run above and the longest run below a cut chosen to maximize tbis length equals or exceeds s: $P(s/s, \text{any cut})$ witb meaning similar to that for $P(s/s, \text{median})$ but for the case where the cut has been chosen to maximize the shorter of the longest runs on each side.

The use of these distributions can be illustrated by tbe calculation of the various run length statistics for a specific example. The 56 speedometer readings presented in Table V and Fig. 5 were observed at one minute intervals during a driver's first trip on a toll highway with

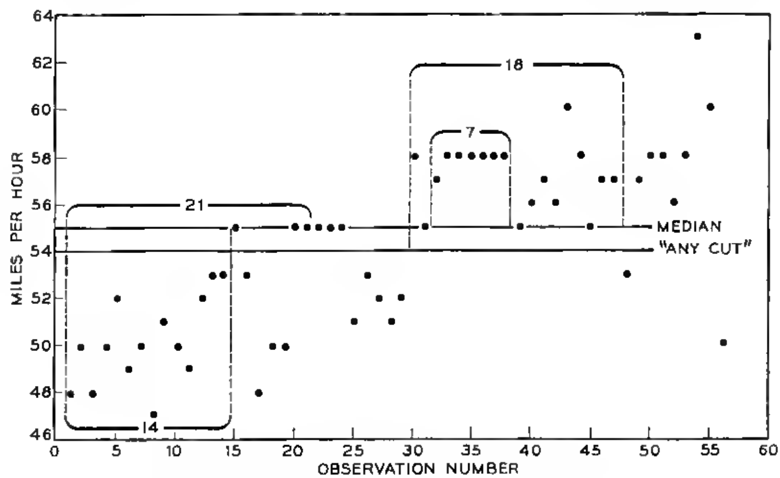

Fig. 5 — Readings at one-minute intervals.

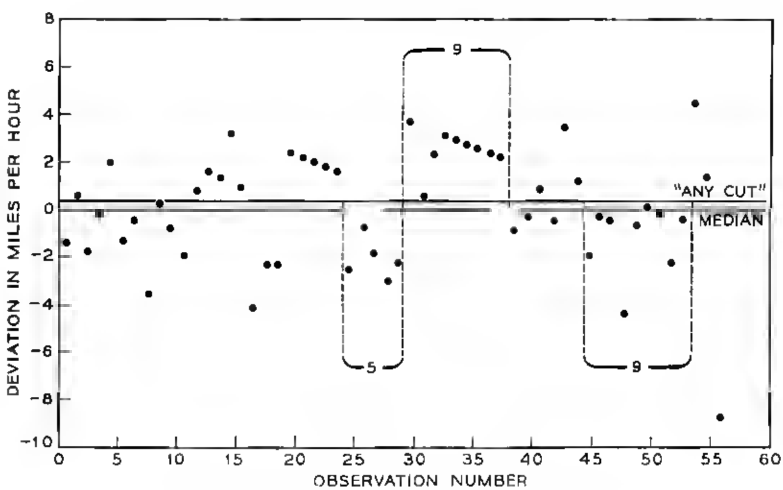

Fig. 6 — Chart for deviations from trend line.

separate traffic lanes. In this instance, nine observations occur at the median (55) with 22 above and 25 below. This is not unusual in experimental work where ties are likely at or near the median. (It should be pointed out that the occurrence of ties makes this a difficult example. Later, this example will be modified by removing a trend and then it will be simpler. Consideration will first be given to runs with respect to the median and then to "any cut.") Various methods of resolving such ties are possible. The most conservative is to use a tied median to terminate a run. The least conservative is to use the tied median or medians for inclusion in the run. Intermediate between these is to consider all possible allocations and their effects on run length. Here, in order to obtain 28 above and 28 below the median, it is necessary to allocate the nine tied at the median so that 6 will be above and 3 below. The run length associated with each such combination would then be obtained and, if desired, the average computed. In this case, the lengths of the various runs obtained by these three methods are as follows:

Type of Run	Run Lengths, s			Limit for $P \leqq 0.01$	Per Cent Below Limit
	Most Conservative	Average	Least Conservative		
Above	7	13.7	18	11 (Table I)	33
Below	14	15.8	21	11 (Table I)	0
Each Side	7	12.8	18	8 (Table II)	1
Either Side	14	16.6	21	11 (Table III)	0
Each Side, Any Cut	14	—	—	9 (Table IV)	—

It will be observed that only one answer results for the "each side of any cut." Also, three of the five tests on the most conservative basis are above their respective limits for a P of 0.01 and all on the other bases. This happens quite frequently in engineering problems.

It is apparent, however, in this case, that there is a consistent trend throughout the set of data. In Fig. 6, this has been removed and the median lies between the 28 points above and the 28 points below. The following statistics are obtained:

Type of Run	Run Lengths, s		P for Observed Run
	Observed Run	Limit for $P \leqq 0.01$	
Above	9	11 (Table I)	0.03
Below	5	11 (Table I)	0.60
Each Side	5	8 (Table II)	0.42
Either Side	9	11 (Table III)	0.05
Each Side, Any Cut	9	9 (Table IV)	0.008

In this case, only the statistic for the longest run on each side of any cut has a P as low as 0.0I. Two others were of the order of 0.05 and the remaining two near 0.50. The explanation of the indicated nonrandomness was identified with human behavior under conditions of learning.

This example raises a question about the treatment of odd sized samples, where the median is a single observation. These may all be reduced to even sized samples by omitting the median. This is unnecessary in the case of the longest run on "each side of any cut" where the P values for a given s for the odd sized sample lie between those for the adjacent even sized samples.

IV. SOME SPECIFIC SAMPLES

Table VI presents the values of probabilities $P(s/-,\ n_1/n_2)$ and $P(-/s,\ n_1/n_2)$, for every possible separation of $10 = n_1 + n_2$ observations into n_1 on one side of a cut and n_2 on the other. Table VII does the same for $20 = n_1 + n_2$ observations. Similarly, the values of $P(s/s,\ n_1/n_2)$ and $P(s/-$ or $-/s,\ n_1/n_2)$ are given in Tables VIII and IX, and Tables X, and XI, respectively.

In Tables XII, XIII, and XIV, the table presented by Mosteller[4] for the three kinds of runs with respect to the median, that is, where $n_1 = n_2$, has been extended to include samples of 60, 100, and 200.

The values of $P(s/s$, any cut) for $n = 10$, 20, 40, and I00 are given in Table XV. It will be noted that the values of the probabilities in this table differ only slightly from those in Table XIII for $P(s/s$, median). For large sample sizes, other considerations suggest that the s-values

Table VI

Probability of an arrangment with a run of length at least s on "one side" of a demarcation value for $n_1 + n_2 = 10$ calculated from equation (1) or (2).

Length of Run s	Total on the "one side", i.e., n_1 or n_2								
	9	8	7	6	5	4	3	2	1
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	0.976	0.833	0.533	0.200	
3	1.000	1.000	0.967	0.786	0.500	0.233	0.067		
4	1.000	0.933	0.667	0.357	0.143	0.033			
5	1.000	0.667	0.333	0.119	0.024				
6	0.800	0.400	0.133	0.024					
7	0.600	0.200	0.033						
8	0.400	0.067							
9	0.200								

TABLE VII

Probability of an arrangement with a run of length at least s on "one side" of a demarcation value for $n_1 + n_2 = 20$ calculated from equation (1) or (2).

Length of Run s	Total on the "one side", i.e., n_1 or n_2																		
	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.990	0.956	0.871	0.718	0.509	0.284	0.100	
3	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.988	0.950	0.870	0.742	0.582	0.415	0.260	0.140	0.060	0.016		
4	1.000	1.000	1.000	1.000	0.996	0.971	0.900	0.779	0.622	0.457	0.307	0.187	0.101	0.046	0.017	0.004			
5	1.000	1.000	1.000	0.986	0.920	0.790	0.621	0.447	0.295	0.178	0.098	0.047	0.019	0.006	0.001				
6	1.000	1.000	0.982	0.889	0.721	0.527	0.351	0.214	0.119	0.060	0.026	0.009	0.005	0.000					
7	1.000	0.995	0.898	0.707	0.492	0.309	0.177	0.092	0.043	0.017	0.006	0.001	0.000						
8	1.000	0.947	0.751	0.509	0.307	0.167	0.082	0.035	0.013	0.004	0.001	0.000							
9	1.000	0.853	0.579	0.341	0.179	0.083	0.034	0.012	0.003	0.001	0.000								
10	1.000	0.711	0.421	0.217	0.098	0.038	0.012	0.003	0.001	0.000									
11	0.900	0.568	0.295	0.130	0.049	0.015	0.004	0.001	0.000										
12	0.800	0.442	0.196	0.072	0.022	0.005	0.001	0.000											
13	0.700	0.332	0.123	0.036	0.008	0.001	0.000												
14	0.600	0.237	0.070	0.015	0.002	0.000													
15	0.500	0.152	0.035	0.005	0.000														
16	0.400	0.095	0.014	0.001															
17	0.300	0.047	0.004																
18	0.200	0.016																	
19	0.100																		

Table VIII

Probability of an arrangement with a run of length at least s on "each side" of a demarcation value for $n_1 + n_2 = 10$ calculated from equation (4).

Length of Run s	n_1 or n_2 n_2 or n_1	1 9	2 8	3 7	4 6	5 5
1		1.000	1.000	1.000	1.000	1.000
2			0.200	0.533	0.833	0.960
3				0.067	0.224	0.333
4					0.029	0.056
5						0.008

Table IX

Probability of an arrangement with a run of length at least s on "each side" of a demarcation value for $n_1 + n_2 = 20$ calculated from equation (4).

Length of Run s	n_1 or n_2 n_2 or n_1	1 19	2 18	3 17	4 16	5 15	6 14	7 13	8 12	9 11	10 10
1		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2			0.100	0.284	0.509	0.718	0.871	0.958	0.990	0.999	1.000
3				0.016	0.060	0.140	0.260	0.413	0.581	0.727	0.784
4					0.004	0.017	0.046	0.100	0.179	0.245	0.274
5						0.001	0.006	0.012	0.042	0.056	0.064
6							0.000	0.002	0.007	0.011	0.013
7								0.000	0.001	0.002	0.002
8									0.000	0.000	0.000
9										0.000	0.000
10											0.000

Table X

Probability of an arrangement with a run of length at least s on "either side" of a demarcation value for $n_1 + n_2 = 10$ calculated from equation (5).

Length of Run s	n_1 or n_2 n_2 or n_1	1 9	2 8	3 7	4 6	5 5
1		1.000	1.000	1.000	1.000	1.000
2		1.000	1.000	1.000	1.000	0.992
3		1.000	1.000	0.967	0.795	0.667
4		1.000	0.933	0.667	0.362	0.230
5		1.000	0.667	0.333	0.119	0.040
6		0.800	0.400	0.133	0.024	
7		0.600	0.200	0.033		
8		0.400	0.067			
9		0.200				

Table XI

Probability of an arrangement with a run of length at least s on "either side" of a demarcation value for $n_1 + n_2 = 20$ calculated from equation (5).

Length of Run s	n_1 or n_2 / n_2 or n_1	1 / 19	2 / 18	3 / 17	4 / 16	5 / 15	6 / 14	7 / 13	8 / 12	9 / 11	10 / 10
1		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3		1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.989	0.966	0.956
4		1.000	1.000	1.000	1.000	0.996	0.971	0.901	0.787	0.684	0.640
5		1.000	1.000	1.000	0.986	0.920	0.790	0.622	0.452	0.337	0.293
6		1.000	1.000	0.982	0.889	0.721	0.527	0.351	0.217	0.134	0.106
7		1.000	0.995	0.898	0.707	0.492	0.309	0.177	0.092	0.046	0.032
8		1.000	0.947	0.751	0.509	0.307	0.167	0.082	0.035	0.014	0.007
9		1.000	0.853	0.579	0.341	0.179	0.083	0.034	0.012	0.003	0.001
10		1.000	0.711	0.421	0.217	0.098	0.038	0.012	0.005	0.001	0.000
11		0.900	0.568	0.295	0.130	0.049	0.015	0.004	0.001	0.000	
12		0.800	0.442	0.196	0.072	0.022	0.005	0.001	0.000		
13		0.700	0.332	0.125	0.036	0.008	0.001	0.000			
14		0.600	0.237	0.070	0.015	0.002	0.000				
15		0.500	0.158	0.035	0.005	0.000					
16		0.400	0.095	0.014	0.001						
17		0.300	0.047	0.004							
18		0.200	0.016								
19		0.100									

Table XII

Probability of an arrangement with a run of length at least s on "one side" of median calculated from equation (1) or (2).

Length of Run s	Sample size, n					
	10	20	40	60	100	200
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	0.97619	0.99994	1.00000	1.00000	1.00000	1.00000
3	0.50000	0.86973	0.99225	0.99956	1.00000	1.00000
4	0.14286	0.45713	0.79885	0.92695	0.99049	0.99994
5	0.02381	0.17849	0.44954	0.63645	0.84289	0.98093
6		0.05960	0.20733	0.33935	0.54439	0.82160
7		0.01703	0.08697	0.15952	0.29185	0.54174
8		0.00395	0.03438	0.07046	0.14251	0.30295
9		0.00065	0.01290	0.02996	0.06642	0.15529
10		0.00006	0.00458	0.01235	0.03015	0.07621
11			0.00153	0.00494	0.01344	0.03656
12			0.00047	0.00192	0.00589	0.01731
13			0.00014	0.00072	0.00255	0.00813
14			0.00004	0.00026	0.00108	0.00378
15			0.00001	0.00009	0.00045	0.00175
16			0.00000	0.00003	0.00019	0.00080
17			0.00000	0.00001	0.00008	0.00037
18			0.00000	0.00000	0.00003	0.00017
19			0.00000	0.00000	0.00001	0.00007
20			0.00000	0.00000	0.00000	0.00003
21				0.00000	0.00000	0.00001
22 or over				0.00000	0.00000	0.00000

TABLE XIII

Probability of an arrangement with a run of length at least s on "each side" of median calculated from equation (6) or (4).

Length of Run s	Sample Size, n					
	10	20	40	60	100	200
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	0.96032	0.99989	1.00000	1.00000	1.00000	1.00000
3	0.33333	0.78582	0.98519	0.99912	1.00000	1.00000
4	0.05556	0.27412	0.66809	0.88729	0.98159	0.99987
5	0.00794	0.06356	0.24933	0.44250	0.72496	0.96284
6		0.01288	0.06820	0.14723	0.33308	0.68619
7		0.00249	0.01647	0.03992	0.10591	0.31377
8		0.00045	0.00379	0.00992	0.02919	0.10573
9		0.00008	0.00085	0.00238	0.00747	0.03027
10		0.00001	0.00019	0.00056	0.00185	0.00800
11			0.00004	0.00013	0.00045	0.00203
12			0.00001	0.00003	0.00011	0.00051
13			0.00000	0.00000	0.00002	0.00013
14			0.00000	0.00000	0.00000	0.00003
15			0.00000	0.00000	0.00000	0.00001
16 or over			0.00000	0.00000	0.00000	0.00000

TABLE XIV

Probability of an arrangement with a run of length at least s on "either side" of median calculated from equation (5).

Length of Run s	Sample Size, n					
	10	20	40	60	100	200
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	0.99206	0.99999	1.00000	1.00000	1.00000	1.00000
3	0.66667	0.95564	0.99931	1.00000	1.00000	1.00000
4	0.23016	0.64014	0.92961	0.98660	0.99938	1.00000
5	0.03968	0.29342	0.64975	0.83041	0.96082	0.99901
6		0.10632	0.34646	0.53147	0.75569	0.95701
7		0.03157	0.15747	0.27911	0.47779	0.76970
8		0.00741	0.06497	0.13100	0.25582	0.50017
9		0.00122	0.02495	0.05754	0.12538	0.28031
10		0.00011	0.00897	0.02414	0.05846	0.14443
11			0.00302	0.00975	0.02642	0.07108
12			0.00093	0.00380	0.01168	0.03411
13			0.00028	0.00144	0.00506	0.01613
14			0.00008	0.00052	0.00216	0.00753
15			0.00002	0.00018	0.00090	0.00349
16			0.00000	0.00006	0.00038	0.00160
17			0.00000	0.00002	0.00016	0.00074
18			0.00000	0.00000	0.00006	0.00034
19			0.00000	0.00000	0.00002	0.00014
20			0.00000	0.00000	0.00000	0.00006
21				0.00000	0.00000	0.00002
22 or over				0.00000	0.00000	0.00000

TABLE XV

Probability of an arrangement with a run of length at least s on "each side" of at least one of all possible demarcation values calculated from equation (22).

Length of Run s	Sample Size, n			
	10	20	40	100
1	1.00000	1.00000	1.00000	1.0000
2	0.97937	0.99997	1.00000	1.0000
3	0.46190	0.89748	0.99713	1.0000
4	0.08413	0.44121	0.83760	0.9986
5	0.00794	0.12994	0.43401	0.9125
6		0.02943	0.15840	(0.5863)*
7		0.00559	0.04544	(0.2561)*
8		0.00093	0.01179	0.0876
9		0.00013	0.00277	0.0263
10		0.00001	0.00066	0.0073
11			0.00015	(0.0020)*
12			0.00003	(0.0005)*
13			0.00001	(0.0001)*
14 or over			0.00000	(0.0000)*

* Values in parentheses were interpolated or extrapolated.

will increase by unity. All this is in accord with the experience of the engineer who did not hesitate to use available information for $P(s/s$ median) as being a good first approximation to $P(s/s$, any cut).

V. SAMPLE ARRANGEMENT DISTRIBUTIONS WITH RUNS OF LENGTH AT LEAST s ABOVE AND BELOW ANY SELECTED CUT

Assume a finite sample of $n = n_1 + n_2$ numbers, of which n_1 have the common property of being above the selected cut and, similarly, n_2 are below. Clearly, the n_2 numbers may be considered as providing $(n_2 + 1)$ cells or partitions of the n_1 numbers above. Some of these cells or partitions will, of course, be empty, particularly when n_1 is less than $(n_2 + 1)$. If at least s of the n_1 numbers are to be in one partition, it would first appear that the number of ways would be proportional to the number of possible partitions, $n_2 + 1$, and also to the number of ways in which the partition boundary points, n_2, may be selected from the remaining numbers, $n - s$, i.e., the combination of $(n - s)$ things taken n_2 at a time. This, however, gives an over-estimate because it counts twice each arrangement that has two partitions of s each, three times for each arrangement that has three partitions of s each, etc. Taking these factors into account, it is found that the number of ways of partitioning the n_1 numbers by means of the n_2 numbers so as to obtain one

or more partitions that contain s or more elements is:

$$\sum_{j=1}^{\left[\frac{n_1}{s}\right]} (-1)^{j+1} \binom{n_2+1}{j} \binom{n-js}{n_2}.$$

Having this, we may write down the probability of an arrangement of n numbers that will contain at least one run of length s or more among the n_1 numbers that are *above* our demarcation value by dividing by $\binom{n}{n_2}$:

$$P(s/-, n_1/n_2) = \frac{1}{\binom{n}{n_2}} \sum_{j=1}^{\left[\frac{n_1}{s}\right]} (-1)^{j+1} \binom{n_2+1}{j} \binom{n-js}{n_2}. \quad (1)^*$$

In a similar manner, we may, by interchanging n_1 and n_2, write down the probability of an arrangement of n numbers that will contain at least one run of length s or more among the n_2 numbers that are *below* our demarcation value:

$$P(-/s, n_1/n_2) = \frac{1}{\binom{n}{n_1}} \sum_{j=1}^{\left[\frac{n_2}{s}\right]} (-1)^{j+1} \binom{n_1+1}{j} \binom{n-js}{n_1}. \quad (2)$$

To assist in determining the probability that an arrangement will contain at least one run of length s or more *on each side* of the demarcation value, let us assume that we have partitioned the n_1 numbers above into r runs of which at least one is of length at least s. These r runs may be associated with $(r - 1)$ runs or partitions of the n_2 in only one way, with $(r + 1)$ runs of the n_2 in only one way, but with r runs of the n_2 in two ways. Each of these sets of possible runs must contain at least one run of length s or more. The resulting partitioning count for s, n_1, and r is:

* Some readers may wish to note that

$$\binom{n}{n_1} P(s/-, n_1/n_2)$$

is the coefficient of x^{n_1} in the expansion of $(1 + x + x^2 + \cdots)^{n_2+1} - (1 + x + x^2 + \cdots + x^{s-1})^{n_2+1}$

$$B(n_i, r) = \sum_{j=1}^{\left[\frac{n_i - r}{s-1}\right]} (-1)^{j+1} \binom{r}{j} \binom{(n_1 - 1) - j(s-1)}{r-1} \text{ for } i = 1, 2 \quad (3)^*$$

and $B(n_2, r - 1)$ and $B(n_2, r + 1)$ are obtained by substituting $(r - 1)$ and $(r + 1)$ respectively for r in (3). All that is needed to secure the desired probability is to find the count of the possible arrangements in both n_1 and n_2 corresponding to each r, sum with respect to r and divide by the total possible arrangements:

$$P(s/s, n_1/n_2) = \frac{1}{\binom{n}{n_2}} \sum_{r=1}^{n_1-s+1} B(n_1, r)[B(n_2, r-1) + 2B(n_2, r) + B(n_2, r+1)]. \quad (4)$$

To find the probability that an arrangement will contain at least one run of length s or more *on either side* of the demarcation value, it should be noted that (4) is counted in both (2) and (1). Thus, this probability is simply:

$$P(s/- \text{ or } -/s, n_1/n_2) = P(s/-, n_1/n_2) + P(-/s, n_1/n_2) - P(s/s, n_1/n_2) \quad (5)$$

where the probabilities on the right hand side of (5) are given by (1), (2), and (4) respectively.

When the median is used as the demarcation value, $n_1 = n_2$, so that $P(s/-, \text{median}) = P(-/s, \text{median})$. In addition, by rearranging terms, $P(s/s, \text{median})$ may be written in the simplified form:

$$P(s/s, \text{median}) = \frac{1}{\binom{2n_1}{n_1}} \sum_{r=0}^{n_1-s+1} [B(n_1, r) + B(n_1, r+1)]^2 \quad (6)$$

where $B(n_1, r)$ and $B(n_1, r + 1)$ are defined by (3) as before. Equation (6) has been used for the new calculations reported here. (See Section IV)

VI. SAMPLE ARRANGEMENT DISTRIBUTIONS FOR RUNS OF LENGTH s OR MORE ON EACH SIDE OF AT LEAST ONE OF ALL POSSIBLE DEMARCATION VALUES

When this derivation was first discussed with a mathematical statistician, he questioned whether anyone would want a criterion based on

* Here, $B(n_i, r) \equiv B(n_i, r, s)$ is the coefficient of x^{n_i} in

$$(x + x^2 + \cdots)^{n_2+1} - (x + x^2 + \cdots + x^{s-1})^{n_2+1}$$

such a distribution. To make it clear that the engineer does want it, assume that we have a set of data and Tables VI to XI, inclusive. The engineer might look for the longest run on either side of the median. Having found it, he might pick a demarcation value that would just include this run. This would give him, for instance,

$$n_2 < \frac{n_1 + n_2}{2}$$

on one side of his demarcation value and

$$n_1 > \frac{n_1 + n_2}{2}$$

on the other side. He might then look for the longest run on the n_1 side. This would give him two long runs that might be equal in length or one shorter than the other. In either case, he could obtain a value of s for the length of run that is equalled or exceeded on each side of his demarcation value. If his total sample happened to be 20, he could obtain $P(s/s, n_1/n_2)$ from (4) or Table IX for n_1, n_2, and s. This probability, however, is based on his having chosen n_1 and n_2 before the experiment and therefore does not indicate what the true probability associated with this process is. At the same time, it is reasonably certain that this is a procedure that many engineers would be inclined to follow if they did not have prior knowledge concerning where to set the demarcation value.

To facilitate the solution, it will be assumed that no two of the n values in a sample of size n are identical. For the analysis given here, n is taken to be even. Study of small samples shows that when n is odd, $P(s, n - 1) \leqq P(s, n) \leqq P(s, n + 1)$. Taking (6) (with the median as initial cut) as a starting point, assume that the demarcation value is moved so that $(n_1 + 1)$ values are on one side and $(n_1 - 1)$ values on the other. This adds a fraction of the total arrangements with runs of length s or more on each side of the new demarcation value equal to:

$$\begin{aligned} &\Delta_1 P(s/s, n_1 + 1/n_1 - 1) \\ &\quad = \frac{1}{(n_1 + 1)\dbinom{2n_1}{n_1 - 1}} \sum_{r=1}^{n_1 - s} B(n_1 - 1, r)[A(n_1 + 1, r - 1) \\ &\qquad\qquad + 2A(n_1 + 1, r) + A(n_1 + 1, r + 1)] \end{aligned} \tag{7}$$

where $B(n_1 - 1, r)$ is given by (3) above and

$$\begin{aligned} A(n_1 + 1, r) = r \sum_{j=0}^{r-1} (-1)^j [r + s - (n_1 + 1) \\ + (j + 1)(s - 1)] \binom{r-1}{j} \\ \cdot \left[\binom{n_1 + 1 - s - j(s-1)}{r-1} \right. \\ \left. - \binom{n_1 + 1 - 2s - j(s-1)}{r-1} \right] \\ + r(r-1) \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \\ \cdot \left[\binom{n_1 + 1 - s - j(s-1)}{r} \right. \\ \left. - \binom{n_1 + 1 - 2s - j(s-1)}{r} \right]. \end{aligned} \tag{8}$$

The essential points in the derivation of (7) and (8) may be perceived most easily by considering some typical computations. Suppose that we wish to derive $\Delta_1 P(4/4, 6/4)$, having previously derived all of the values of $P(s/s, 5/5)$ from (6). The possible combinations with a run of at least 4 on each side of a cut with 6 above and 4 below have the following orders:

1. 6 above and 4 below, or 4 below and 6 above,
2. 5 above, 4 below, and 1 above, or 1 above, 4 below, and 5 above, and
3. 4 above, 4 below, and 2 above, or 2 above, 4 below and 4 above.

The simplest of these is the first. Starting with the value of $P(4/4, 5/5)$ as given by (6), we now wish to determine how much additional probability is associated with moving the cut from the median to a point where 6 are above and 4 are below. Since there are 6 possible locations in the new arrangement for the value that was moved from below to above the cut and $\binom{10}{6}$ ways for arranging 6 above and 4 below, the total possible combinations of these provides the factor given in the denominator of (7), in this case $6\binom{10}{6}$. Since there is only one combination possible for 4 items taken 4 at a time, $B(4, 1)$ as given by (3) is as might be expected unity. Then, since we must have at least one run above the cut, $A(6, 0)$ must be 0. The first important question relates to the value of $A(6, 1)$. Since there is only one run of 6, it is easy to see that a run of length 4 or more must have occurred above the median if the value

moved from below is now in position 1, 2, 5, or 6 in the new run. Hence, there are only two possible locations for the value moved that give new combinations that have not been counted with respect to the median. At this point, it will be observed that for this case, $r = 1$, this value is given by $2s - (n_1 + 1)$. Since there is exactly one run on each side of the new cut, the coefficient 2 appears before the $A(n_1 + 1, r)$ in (7) to take account of the two ways that these runs may be arranged, namely, 6 above followed by 4 below and 4 below followed by 6 above.

Now consider the ways in which we may have two runs with the restriction that one must be of length 4 or more. This is to be given by $A(6, 2)$. In this case, there are two such run combinations, one with runs of lengths 5 and 1, and one with runs of lengths 4 and 2. Obviously, the value that was moved could not have been in the short run in either case because these arrangements would have had long runs of length 4 or more that would have been counted with respect to the median. In the case of the run of length 5, it could not be on either end but in the run of length 4, it could be at any one of the positions in the run. We also observe that with two runs of dissimilar lengths, the positions of the runs may be interchanged. This gives in this case a factor 2. Hence, we find that $A(6, 2)$ is $2 \cdot 3 + 2 \cdot 4$, or 14. To conform with (8), this sum would have to be written as $2 \cdot 3 \cdot 2 + 2 \cdot 1$, although, at this point, it may not be clear that this is a reasonable thing to do. However, by extending the investigation step by step, it is found that the various terms in (8) are required. Specifically, the j becomes necessary when $n_1 + 1$ becomes greater than $2s - 1$ and the binomial coefficients with terms in $2s$ are introduced so that any combination that already has a run of length s on the basis of the median will not be counted again.

Obviously, this process may be continued by moving the cut to include $(n_1 + 2)$ values on one side and leave $(n_1 - 2)$ values on the other. Proceeding in this way, the fraction added in going from $(n_1 + i - 1)$ values above and $(n_1 - i + 1)$ values below to $(n_1 + i)$ above and $(n_1 - i)$ below is given by:

$$\Delta_i P(s/s, n_1 + i/n_1 - i)$$

$$= \frac{1}{(n_1 + i)\binom{2n_1}{n_1 - i}} \sum_{r=1}^{(n_1-i)-(s-1)} B(n_1 - i, r) \tag{9}$$

$$\cdot [A(n_1 + i, r - 1) + 2A(n_1 + i, r) + A(n_1 + i, r + 1)]$$

where $B(n_1 - i, r)$ is defined by (3) above and

$$
\begin{aligned}
A(n_1 + i, r) = r \sum_{j=0}^{r-1} (-1)^j [r + s - (n_1 + i) + (j + 1)(s - 1)] \\
\cdot \binom{r-1}{j} \left[\binom{n_1 + i - s - j(s-1)}{r-1} \right. \\
\left. - \binom{n_1 + i - 2s - j(s-1)}{r-1} \right] + r(r-1) \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \\
\cdot \left[\binom{n_1 + i - s - j(s-1)}{r} - \binom{n_1 - i - 2s - j(s-1)}{r} \right].
\end{aligned} \tag{10}
$$

One of each of these Δ's is added in going from the median to each side. Therefore, the desired probability of an arrangement with runs of length s or more on each side of at least one of all possible demarcation values is:

$$P(s/s, \text{any cut}) = P(s/s, n_1/n_1) + 2 \sum_{i=1}^{n_1 - s} \Delta_1 P(s/s, n_1 + i/n_1 - i). \tag{11}$$

VII. ASYMPTOTIC DISTRIBUTIONS

Intuitively, the asymptotic distribution of arrangements with 0, 1, 2 etc., runs of length s or more for $n_1/n = e_1$, a constant, would be expected to become Poisson Exponential as n becomes large. Referring to Mood,[3] the expected number of runs of length s or more on one side of a demarcation value is his expression (3.13), which may be written:

$$E(r_{1s}) = (n_2 + 1) \frac{n_1^{(s)}}{n^{(s)}} \sim n e_1^s e_2 \qquad \text{for } n_1 \text{ and } n_2 \text{ large} \tag{12}$$

where $E(r_{1s})$ is the expected number of runs of length s or more on the side of the cut designated 1; superscript (s) designates a factorial moment, e.g.,

$$n^{(s)} = n(n-1)(n-2) \cdots (n - s + 1) \tag{13}$$

and e_1 and e_2 are written for n_1/n and n_2/n, respectively.

The variance is his expression (3.15), or

$$
\begin{aligned}
\bar{\sigma}_{r_{1s} r_{1s}} = \bar{\sigma}_{ss} = \frac{(n_2 + 1)^{(2)} n_1^{(2s)}}{n^{(2s)}} + (n_2 + 1) \frac{n_1^{(s)}}{n^{(s)}} \\
\cdot \left(1 - (n_2 + 1) \frac{n_1^{(s)}}{n^{(s)}}\right)
\end{aligned} \tag{14}
$$

$$\sim n e_1^s e_2 (1 - s^2 e_1^{s-1} e_2^2 - e_1^s)$$

$$\sim n e_1^s e_2 = E(r_{1s}) \quad \text{for } s, n_1, \text{ and } n_2 \text{ large.}$$

Corresponding expressions for the side designated 2 may be obtained by interchanging the subscripts, 1 and 2, in equations (12) and (14).

Mood[3] also derives an expression (3.18) for the covariance of numbers of runs equal to or greater than specified lengths on the two sides of the demarcation value. For runs of length s or more on each side, this becomes:

$$\begin{aligned}\bar{\sigma}_{r_{1s}r_{2s}} &= \frac{n_1^{(s+1)}n_2^{(s+1)}}{n^{(2s)}} + \frac{2n_1^{(s)}n_2^{(s)}}{n^{(2s-1)}} - \frac{(n_1+1)(n_2+1)n_1^{(s)}n_2^{(s)}}{n^{(s)}n^{(s)}} \\ &\sim n\, e_1^s e_2^s(s^2 e_1 e_2 - s + 1) \\ &\sim n\, s^2 e_1^{s+1} e_2^{s+1} \text{ for } s,\ n_1,\text{ and } n_2 \text{ large.}\end{aligned} \tag{15}$$

From (14) and (15), it is clear that the covariance between long runs on the two sides becomes negligible for s, n_1, and n_2 large and the occurrence of long runs on each side may be treated as independent.

Since Mood[3] has shown (his Theorem I) that the distribution of the number of runs of length s or more on one side is asymptotically normal and by (12) and (14) above, the first two moments are those of a Poisson Exponential, the asymptotic probabilities of arrangements with runs of length s or more may be approximated by:

On side 1:

$$P(s/-,\ n_1/n_2) \doteq 1 - e^{-ne_2e_1^s}; \tag{16}$$

On side 2:

$$P(-/s,\ n_1/n_2) \doteq 1 - e^{-ne_1e_2^s}; \tag{17}$$

On each side: ·

$$P(s/s,\ n_1/n_2) \doteq (1 - e^{-ne_2e_1^s})(1 - e^{-ne_1e_2^s}); \tag{18}$$

On either side:

$$P(s/- \text{ or } -/s,\ n_1/n_2) \doteq 1 - e^{-ne_1e_2(e_1^{s-1}+e_2^{s-1})}. \tag{19}$$

When the median is being used as the demarcation value, that is, when $e_1 = e_2$, these become:

On side 1 or on side 2 alone:

$$P(s/-,\text{ median}) = P(-/s,\text{ median}) \doteq 1 - e^{-n\cdot 2^{-(s+1)}}; \tag{20}$$

On each side:

$$P(s/s,\text{ median}) \doteq (1 - e^{-n\cdot 2^{-(s+1)}})^2; \tag{21}$$

On either side:

$$P(s/- \text{ or } -/s,\text{ median}) \doteq 1 - e^{-n\cdot 2^{-s}}. \tag{22}$$

Asymptotic relationships of this type do not add much to the solution of the practical problem of calculating probabilities associated with samples of 100 or less. They do, however, suggest that doubling sample sizes for a given probability should increase s by unity. This is in close agreement with the calculations for finite sample sizes. This observation suggested the treatment in the next section.

VIII. RELATIONSHIPS BETWEEN s AND n FOR CONSTANT PROBABILITY

From (20), (21), and (22), it is clear that, for constant probability, s is asymptotically a simple function of n for each of the arrangement distributions considered for runs relative to the median. Specifically, we obtain:

On side 1 or on side 2:

$$s = \frac{\log n - \log(-\log_e(1 - P))}{\log 2} - 1 \qquad \text{where} \quad P = P(s/-, \ \text{median}) = P(-/s, \ \text{median}); \tag{23}$$

On each side:

$$s = \frac{\log n - \log(-\log_e(1 - \sqrt{P}))}{\log 2} - 1 \qquad \text{where} \quad P = P(s/s, \ \text{median}); \tag{24}$$

On either side:

$$s = \frac{\log n - \log(-\log_e(1 - P))}{\log 2} \qquad \text{where} \quad P = P(s/- \text{ or } -/s, \text{median}). \tag{25}$$

After considering equations (23), (24), and (25), it is quite obvious that an equation similar to (24) in the same way that (25) is similar to (23) could be written, i.e.;

$$s = \frac{\log n - \log(-\log_e(1 - \sqrt{P}))}{\log 2} \qquad \text{where } P = P[(s/- \text{ or } -/s)/(s/- \text{ or } -/s), \text{median}] \tag{26}$$

but what is the meaning of P? It is clear that the P in (26) is approximately the square of the P in (25). So far, however, no analytic justification for (26) has been obtained, although the P in (26) is obviously

Table XVI

Constants for equation (27) calculated from equations (23) to (26) and tables VII to X

Equation	Table	P	A	B	C	Differences at n equal to					
						10	20	40	60	100	200
23	VII	0.001	5.151	126.6	−266.5	0	0	0	0	−0.01	+0.01
		0.01	4.863	53.19	−105.1	0	+0.02	−0.02	−0.01	0	+0.02
		0.02	4.445	39.61	−79.16	0	+0.01	−0.01	−0.01	0	+0.02
		0.025	4.306	35.34	−71.08	0	+0.01	−0.01	−0.01	0	+0.02
		0.05	3.127	28.03	−57.71	0	+0.01	−0.01	−0.01	0	+0.02
		0.10	3.127	13.95	−32.83	0	+0.01	−0.01	−0.01	0	+0.01
		0.50	0.5297	−3.126	−0.1306	0	0	0	0	0	0
		0.90	−2.576	−6.757	10.96	0	0	0	+0.01	0	0
		0.95	−3.442	−7.244	13.84	0	0	0	0	0	0
		0.975	−4.227	−7.185	15.44	0	0	0	0	0	−0.01
		0.98	−4.441	−7.326	16.29	0	−0.01	0	0	0	−0.01
		0.99	−5.128	−7.164	17.78	0	−0.01	+0.01	0	0	−0.01
		0.999	−7.829	0.3596	8.278	0	0	−0.01	−0.01	+0.06	−0.04
24	VIII	0.001	−0.3002	23.36	−32.54	0	−0.01	+0.02	0	0	−0.01
		0.01	0.0467	10.93	−13.30	0	0	0	0	0	0
		0.02	0.0048	8.596	−11.14	0	0	0	0	0	0
		0.025	−0.0005	7.612	−9.868	0	0	0	0	0	0
		0.05	−0.0672	4.695	−6.288	0	0	0	0	0	0
		0.10	−0.2136	1.668	−2.139	0	0	0	0	0	0
		0.50	−1.573	−4.142	6.576	0	0	0	+0.01	0	−0.01
		0.90	−3.660	−6.518	13.04	0	0	0	0	0	−0.01
		0.95	−4.408	−6.591	14.75	0	0	0	0	0	−0.01
		0.975	−5.039	−6.665	16.21	0	0	0	0	0	−0.01
		0.98	−5.218	−6.728	16.72	0	0	0	+0.01	0	−0.01
		0.99	−5.822	−7.068	18.19	0	0	0	+0.01	0	−0.01
		0.999	−7.430	−6.861	23.26	0	−0.01	0	+0.01	0	−0.01
25	IX	0.001	4.879	154.6	−324.2	0	0	+0.01	0	−0.02	+0.01
		0.01	4.902	73.86	−145.9	0	+0.01	−0.01	−0.01	0	+0.02
		0.02	4.764	55.83	−109.5	0	+0.01	−0.01	−0.01	0	0
		0.025	4.611	51.88	−102.5	0	+0.01	−0.01	−0.01	0	+0.02
		0.05	4.432	35.72	−71.31	0	+0.02	−0.02	−0.01	0	+0.02
		0.10	3.847	25.64	−54.79	0	+0.01	−0.01	−0.01	0	+0.02
		0.50	2.524	1.680	−13.88	0	+0.01	−0.01	0	0	+0.01
		0.90	1.141	−9.694	7.601	0	0	0	0	0	0
		0.95	0.759	−12.16	12.83	0	0	0	0	0	0
		0.975	0.422	−14.08	17.14	0	0	0	0	0	0
		0.98	0.356	−14.74	18.61	0	0	0	0	0	0
		0.99	0.081	−16.61	22.91	0	0	0	+0.01	0	0
		0.999	−0.600	−21.68	34.97	0	−0.01	+0.02	0	−0.02	+0.01
26	X	0.001	−4.176	60.34	−69.24	+0.01	−0.01	0	—	+0.01	—
		0.01	−2.176	39.32	−49.98	+0.01	+0.01	−0.01	—	0	—
		0.02	−1.762	34.71	−47.65	0	+0.01	−0.02	—	−0.02	—
		0.025	−1.427	32.01	−45.06	0	+0.02	−0.01	—	−0.01	—
		0.05	−0.830	26.26	−41.04	0	+0.02	−0.01	—	−0.02	—
		0.10	−0.356	21.25	−38.10	0	+0.01	−0.01	—	−0.01	—
		0.50	1.069	1.932	−14.67	0	+0.01	−0.01	—	0	—
		0.90	1.136	−10.55	5.014	0	+0.02	−0.02	—	0	—
		0.95	1.369	−17.16	18.39	0	+0.01	−0.02	—	+0.02	—
		0.975	1.222	−19.89	24.12	−0.01	+0.01	−0.02	—	+0.02	—
		0.98	1.007	−19.78	24.69	−0.01	+0.01	−0.02	—	+0.01	—
		0.99	0.679	−21.66	30.15	−0.01	0	−0.02	—	+0.02	—
		0.999	0.408	−29.73	48.79	0	0	−0.02	—	+0.02	—

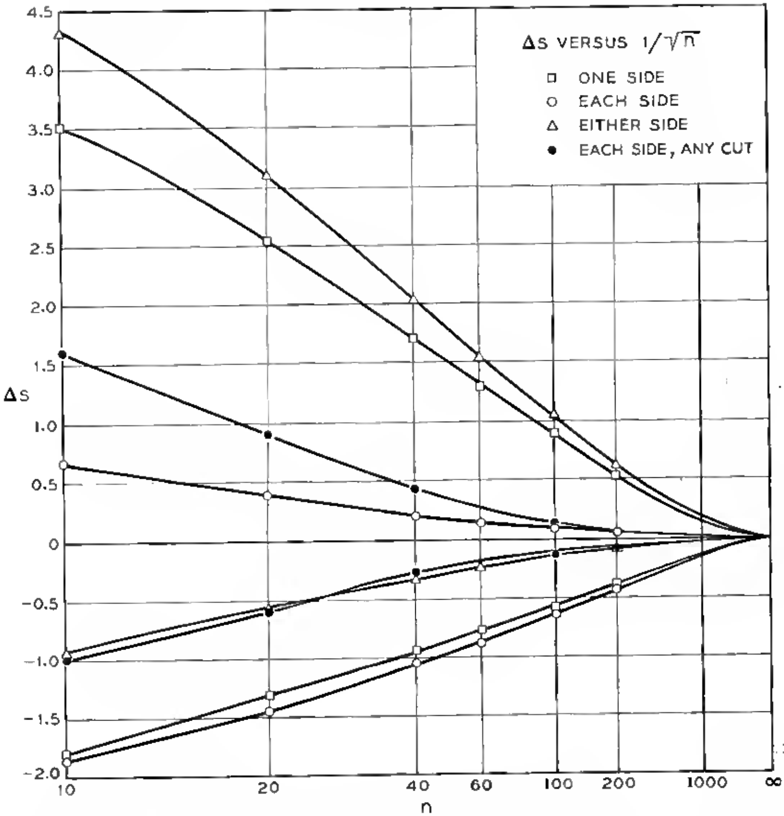

FIG. 7 — Differences between interpolated values of s computed from Tables XII to XV, inclusive, and appropriate equations (23) to (26), inclusive, for $P = 0.01$ and 0.99.

the maximum value possible for $P(s/s$, any cut). Nevertheless, as we shall see below, it appears to predict empirically the large sample behavior of runs above and below any cut even better than (23), (24), and (25) predict the large sample behavior of the other types of run.

For this comparison, values of s corresponding to particular values of P were interpolated (in a few cases, extrapolated) from the exact determinations of Tables XII to XV. Since the distributions for each sample size in these tables had been found to be mildly deviant from log-normal, the interpolation process first obtained a three point log-normal relationship in the P area of interest by changing the s-scale to an $(s + a)$-scale. Here, a is the constant that must be added to s to produce the log-normal relationship in the interval under consideration. Values of s for each P, n, and type of run were obtained to four decimal places.

In each case, the difference between the interpolated value and that given by the appropriate equation (23), (24), (25), or (26) was calculated. At this point, it was found that some of these differences for a particular P and type of run could be approximated by linear equations in $1/n$ or $1/\sqrt{n}$. In view of this, all have been fitted by the equation:

$$\Delta s = \frac{A}{\sqrt{n}} + \frac{B}{n} + \frac{C}{\sqrt{n^3}} \tag{27}$$

The constants, A, B, and C, have been recorded in Table XVI. The agreement between the values given by this equation and the differences on which they were based seldom exceed 0.02. Thus, it was assumed that (27) provided a reasonable approximation for extrapolation to the larger sample sizes for which values are shown in Tables I to IV and in Figs. 1 to 4.

To illustrate the agreement with (27), some typical results for P's of 0.01 and 0.99 are given in Fig. 7. All show that the differences converge in a reasonably uniform manner to zero at infinity.

IX. ACKNOWLEDGEMENTS.

The author would like to thank Miss Dorothy T. Angell for making computations for samples of 60, 100, and 200 given in Tables XII to XIV; Mrs. Helen K. Cunningham for the tedious calculations for the sample of 100 in Table XV; Mrs. Mary Jane Weideli and Mrs. Rita R. Singer for their check of the material in Section VIII; and John W. Tukey for his persistent prodding and encouragement to present this partially incomplete report as well as his comments and suggestions relating to presentation.

REFERENCES

1. W. E. Campbell, Use of Statistical Control in Corrosion and Contact Resistance Studies, Trans. Electrochem. Soc., **81,** 377–390, 1942.
2. W. G. Cochran, An Extension of Gold's Method of Examining the Apparent Persistance of One Type of Weather, Quarterly Journal of the Royal Meteorological Society, **44,** pp. 631–634, 1938.
3. A. M. Mood, The Distribution Theory of Runs, Ann. Math. Stat., **11,** pp. 367–392, 1940.
4. F. Mosteller, Note on an Application of Runs to Quality Control Charts, Ann. Math. Stat., **12,** 228–232, 1941.
5. W. A. Shewhart, Contribution of Statistics to the Science of Engineering, University of Pennsylvania Bicentennial Conference, Volume on Fluid Mechanics and Statistical Methods in Engineering, pp. 97–124, 1941, University of Pennsylvania Press.
6. A. C. Walker and P. S. Olmstead, Textile Yarn Abrasion Test, Textile Research Journal, **15,** pp. 201-222, 1945.
7. P. S. Olmstead, How to Detect the Type of an Assignable Cause, Monograph 2106, Bell Telephone System Technical Publications, and Industrial Quality Control, **9,** pp. 32–34 and 36, Nov., 1952, and pp. 22–32, Jan., 1953.